

SC22/WG20 meeting in Cairo, November 16-20, 1997 Draft Minutes

November 20, 1997

1. Introduction and announcements

Director Abdel Baset El-Sebai (Chair of the Egyptian Organization for Standards EOS) addressed the meeting with encouraging words about the importance of international standards for Egypt and its technical and economic connections with the rest of the world. He wished us success for our work.

The convener welcomed the participants and thanked IBM Egypt for hosting the meeting.

2. Introduction of participants

Name	National Body	Affiliation
Ahmed, Afaf Taka	Egypt	EOS
Kido, Akio	Japan	IBM
Kung, Michael	USA	Oracle
LaBonté, Alain	Canada	Conceil de tresor
Sato, T.K.	Japan	HP
Sherif, Khaled	Egypt	IBM
Simonsen, Keld	Denmark	RAP
Soor, Baldev	Canada	IBM
Winkler, Arnold	convener	Unisys

3. Appointment of chairperson, secretary, and drafting committee

Chair: Winkler
 Secretary: Winkler
 Drafting: Simonsen

4. Approval of prior meeting's minutes [N518]

Minutes approved

5. Future Meeting Schedule and Plans

June 15-19, 1998	Dublin	Ireland (alternate US)
October 19-23, 1998		Israel
June 15-19, 1999	Philadelphia	USA
Fall 1999	Copenhagen	Denmark

Action: confirm with Everson move to June 15-19, 1998

6. Recognition of new documents and assignment to agenda items

542	Final agenda for meeting in Cairo, Nov. 16-20, 97	Winkler	97-11-16	admin
543	List of participants in Cairo, Nov. 16-20, 1997	Winkler	97-11-20	admin
544	Minutes - Cairo, November 1997	Winkler	97-11-20	admin
545	Resolutions - Cairo, November 1997	Winkler	97-11-20	admin
546	Liaison statement from SC22/WG15	Simonsen, WG15	97-10-20	admin
547	Contribution on Cyrillic sorting	Michael Everson	97-11-	22.30.02.02
548	Contribution from WG20 to CAW - draft	SC22/WG20	97-11-15	admin

549	Voting results and comments to CD 14652 ballot	SC22 N2612	97-11-14	22.30.02.03
550	Document register N510 - N550	Winkler	97-11-20	admin

7. Approval of Agenda

Agenda is approved with addition of 16.2 (registration of cultural elements).

8. Liaison Reports

8.1 Additions/deletions/changes to liaisons [SD-4]

8.1.1 SC22/WG4, COBOL

No written report. Ann supports identifiers as in Annex A.

8.1.2 SC22/WG15

546	Liaison statement to SC22/WG15	Simonsen, WG15	97-10-20	admin
-----	--------------------------------	----------------	----------	-------

WG15 is working on guidelines for locales for POSIX. Euro is being addressed as defect report.

8.1.3 SC2/WG2

Kido: due to changes in organizational structure of WG20 and/or WG2. Sato: what about Arnold. Keld: what about Alain. Arnold accepts.

Resolution: Arnold replaces Sato after the Seattle meeting in March 1998.

8.1.4 WG14

No written report. C standard is being revised, date, time, etc... Euro is being addressed. Extended identifier \uxxxx is being introduced.

Action Keld: Distribute draft 11.7 extraction for identifiers to WG20.

Action Keld: inform WG14, WG15 and WG21 about euro issues.

POSIX and C standard should be aligned - proposal from Keld.

8.1.5 WG21

No written report.

Final DIS is being prepared.

Action Keld: inform WG21 about new Annex A after this meeting.

8.1.6 WG5

No written report, support Annex of 10176.

8.1.7 GUIDE/SHARE Europe

Nothing new.

8.1.8 SC18/WG9

changed its name - JTC1 WG5 - user interfaces (Icons, keyboards, symbols, etc...). Modifying keyboard standard to contain euro

Action Winkler: change liaison list, SC18 does not exist any more.

8.1.9 CEN TC 304

No written report. Re-organizing, interested in euro. New stuff: cultural heritage.

8.1.10 TC37

No written report. ISO 12199 sorting standard being published.

8.1.11 ITU-T

No written report.

8.1.12 OpenGroup

Action Keld: ask for name of Liaison to WG20

SQL standard in CD ballot - octet based, binding is not conforming to

9. Review of prior meetings action items [SD-5]

all completed

10. Revision of TR 10176

531	Proposed disposition of comments to TR 10176	Kido/Noda	97-09-25	22.13
532	Final disposition of comments to TR 10176	Kido/Noda		22.13
533	Final text of TR 10176 for TR ballot	Kido/Noda		22.13

Discussion of N531, proposed disposition of comments to TR 10176.

Kido: most controversial is the presence of annex A - no NB contributions have been received, but some contributions from working group convenors. Question is the repertoire of characters for identifiers - minimum (ISO 646) or maximum (all of 19646), or anything in between.

Arnold: make list as wide as possible, let individual standards apply restrictions.

Keld: we should specify minimum and let individual languages add to it. C and C++ are going to allow all of 10646 letters.

Kido: 4 alternatives: minimum set 646 invariant, maximum is everything in 10646, 2 possibilities in between, one with combining characters, one without.

Soor: conformance to a standard should be required, additions are allowed per implementation. Therefore, minimum set must be specified, languages can expand as they want.

Mock ballot: Keld: level 2 OK, 1 list only. Canada: 1 list only, no double coding.

Conclusion: Level 2 of combining characters is accepted by anybody who voted YES without changing their vote.

Additional question: what about digits? Digits should be allowed, some languages not in first position.

Character “underscore” - exclude, also other punctuations and special characters. Compatibility zone characters : half width katakana and double width kanjii remove. Some superscript should be included. Ligatures (presentation forms) to be excluded (Khaled). 309B and 309C remain on the list. Remove combining sound marks, CJK symbols, ideographic iteration marks.

Continuation of discussion on 11/15/97:

Kung: needs Level 3, and combining characters - Kido: not acceptable at this time, will be explained. US will most likely not change their NO vote.

Keld comments: UK issues should be addresses in further revision of 10176. This is a implicit commitment from WG20

Kido: should we have a minimum list and allow languages to expand, or should we have a wide definition with restrictions applied in languages or implementations. Problem is not only in identifier inside a program, but mainly the external (link) identifiers to other program elements. Keld agrees to “modify”.

Plans for next revision: Where will it be done - WG20 or SC22 or else ??

Further progression:

Final disposition of comments	Kido	provide for web publication by 12/1/97
Comments to DoC by 12/8/97, editor and convener action	Kido	automatically approved, if no comments
TR: Check for English	Winkler	end of 1997
TR: Complete editing	Kido	February 1998
TR: Check on the web by WG20	all WG20	March 6, 1998
TR: Send to ITTF for publication	Winkler	March 1998

Action item Kido: check all character names in the disposition of comments, correct as needed, and send new final text to Winkler .

Timeplan agreed for progression of TR - see above.

11. International string ordering ISO/IEC CD 14651

527R	Disposition of comments to CD 14651 ballot in SC22 N2466 (WG20 N521)	Alain LaBonté SC22 N2606	97-07-01	22.30.02.02
537	ISO/IEC 14651 FCD - International string ordering for FCD ballot	Alain LaBonté SC22 N2607	97-11-05	22.30.02.02
538R	Request to national bodies concerning an informative annex to ISO/IEC provisionally removed from FCD 14651 (see in connection with N537)	Alain LaBonté SC22 N2608	97-11-05	22.30.02.02
547	Contribution on Cyrillic sorting	Michael Everson	97-11-	22.30.02.02

The standard is in ballot, no major discussion is possible at this time. Due date: April 30, 1998.

12. Cultural convention specification standard ISO/IEC CD 14652

528	ISO/IEC 14652 CD - Specifications of cultural conventions	Keld Simonsen SC22 N2504	June 1997	22.30.02.03
-----	---	-----------------------------	-----------	-------------

Plan:

Study document, Keld will provide preliminary disposition. Discuss further on Tuesday. Arnold will send an e-mail to SC22 secretariate for final vote count and any more comments. If necessary, teleconference for the disposition of comments. Will be discussed on Tuesday.

Got answer from Bill Rinehuls - no more comments, full vote count

Vote for registration: 9 YES, 3 NO,
CD ballot: 7 YES, 2 YES with comments, 3 NO

Kido: is this a specification method or a compiler definition? Conformance statement is essential. If this is a compiler format, we need specification of locale under which it will be compiled.

Keld: compiler specific parts are not included.

Kido: if included, the standard would have to go back to

Sato: clear specification of FDCC set must be prepared.

Keld presents his draft disposition of comments, based on the comments received from Rinehuls. Concern is that this standard, at least the collation portion, must be ready together with IS 14651.

General issues:

why do we need this standard

conformance

FDCC explanation

Approach: read through, get feel from the WG, then Keld can add text according to direction of WG.

Describe benefits, not history. Users are not only ISO members.

Canada:

Lunar calendar needs to be added, also Coptic, Japanese, ... Text needed from Canada.

Currency: Keld will propose (see DK comment 2.1). Kido requests double currency, with euro as a specific case. Soor describes the impact on a change on the LOCALEDEF compiler, the behaviour of the alternate currency in LC_MONETARY, and strfmon function (3-letter code, symbol). Question: should exchange rate be included? Check with Baldev

Calendary: Canada will provide

LC_TIMEZONE, etc...: discussion about the format of the LC_ statements: either LC_X_EEE... or LC_EEE... . Sato assumes that this difficulty comes from the misunderstanding of FDCC set - he will explain this afternoon. Idea is that LC_X_ are non-standardized definition, LC_ are reserved categories for further standardization. Solution: remove restriction of LC_X_ but make it a recommendation. Possible conflict on name space of LC categories. 6 major categories are mandatory to assure portability, other categories can be introduced on the user's risk that the name can later be used for a different standardized category.

3 - Wasted category : Keld will add additional text as requested by Canada.

5 - Portable character set : must be part of every charmap. Add long name and/or long name to the glyph. Also, point to portable character set.

6 - Printable vs graphical : what about space characters in 10646?? Also, "class" of zero, one, etc in Latin, Arabic, Hindu, Kanjii, etc...Canada wants to be able to use local digits in addition to portable digits. Japan supports this request. For digits: array of 0-9 digits, also for local digits in the same order.

6 - Upper/lower: rejected is OK.

6 - Transformation: must be called "transliteration" in this context.

LC_CTYPE: accepted. Check the completeness.

Keywords: reference to definition is required, (also good definition requested by Japan).

Number of values in LC_COLLATE - POSIX requires 7, Sort only requires 4 for its use. Remove the word default.

Collating Order (4.1.2.5) will be moved before 4.1.1.2

11 - accepted

12 - accepted

13 - 4.1.2.11 will be 4.1.2.10.1

14

15 - will be clarified.

16 - monetary symbol is predefined as comma - this is for Europe only. Proposal is to keep it empty, that allows anything. This is in line with other parameters, such as positive and negative sign, etc..

16 - value -1 is for POSIX compatibility, Canada agrees.

16 - euro handling : new proposal requested

16 - comma or period as decimal point? make it empty string, with a note that no default is applied. Also, make "thousand-separator" to "group-separator" for groupings of 4.

18 - era keyword. Is this keyword the only one that is mandatory. Since all keywords are mandatorily accepted, the word mandatory will be deleted.

18 - number of month - 13 allowed for non Gregorian calendar. No conversion specified.

18 - am/pm keywords: what if unspecified ? show empty string., no default

18 - era keyword: no default to be used

19 - accepted

19 - time zone with Summer time, multiple time zones, ... put on issues list, revisit after requirement statement

Question: what about Arabic countries, where the first day of the week is Saturday (%U) . Written comment with next ballot is invited.

Soor: why do we need a %OW and not %OF

21 - incomplete repertoire map : for compatibility

Danish comments

2.1 - euro and alternate currency.

2.2 - weights for LC_COLLATE - new keyword "weight-equivalence"

2.3 - YES, NO, and CANCEL string - Kido suggests OTHER instead of CANCEL.

2.4 - ISO 2022 support for charmap specification: Kido suggests RFC like method - implementation as example. Rough text for implementation description will suffice for the moment.

Japanese comments:

1 - LC_PAPER : Separate paper size from paper margin. Use millimeters, but recommend an internal conversion from inches.

2 - LC_POSTAL for addresses. Separate names from postal addresses.

2 - LC_TELEPHONE is accepted.

2 - COLOR - use number for color. Also specify local color names (“skin color” in Japan)

Measurement system: ISO system and US systems - needs subdivision. Also needs selection of treated measurements.

Name: First, last, MI, ... Many issues: multiple names, given names, self naming, name changes, title, suffix,

Kido: this standard is not only for JTC1 members - all countries should be invited to comment. Add new categories gradually.

3 - Name and meaning of FDCC-set, FDCC category, etc... Categories need to be conforming individually. This allows to easily add new categories with their own conformance. Winkler: multi-part standard ? Sato: No!

3 - FDCC-set: User categories - LC_X_ with recommended structure. Add conformance per category. Explain FDCC and FDCC-set. (FDCC-set-element == FDCC).

Kido - Sato: POSIX dependence must be re-worded.

J3 - accepted, see also Kido-sans comments

J4 - delete a number of references - move to bibliography

J5 - remove description of collation - accepted.

J6 - J7 - affirmative responses - accepted

J8- accepted formal syntax description

J9 - ellipses

J10 - accepted , remove “normative”

J11 - keywords can be repeated in different categories, do not have to be unique in FDCC-set - accepted

J12 - restructuring accepted

J13 - one repertoire map per FDCC set - needs wording

J14 - charmap: Japan will provide requirement for next ballot

J15 - character representation, accepted

J16 - accepted

J17 - accepted

J18, J19, J20 - accepted

J21 - character string transformation. The word transformation has been changed to transliteration. See Canadian comment #7.

J22 - accepted,

J23 - alpha classification for IPA characters: rejected comment, Japan will come back with decision later.

J24, J25, J26 - accepted

J27 - define “forward” and “backward” in collation related to LTR and RTL languages: accepted

J28 - accepted, elif keyword usage

J29 - add note saying that that is using the common templates in the sort standard

J30 - see Canadian comments 16, accepted

J31 - Days per week: in Japan, there are variable length weeks (5, 7, 12 days). Clear specification is needed. Add also start of week (Egypt Saturday, US Sunday, Europe Monday), first working day of the week (US Monday, Europe Monday, Egypt ??, Saudi Arabia ??), add directionality of the calendar.

J31 - Sequence of day, date, time. Rejected comment based on preference of the editor.

J32 - accepted

J33 - accepted

J34 - accepted

J35 - accepted

J36 - open discussion about the meaning of “byte” and the functionality of the utility. Plan of Keld: define file format in this standard and compiler (localedef) utility in the API standard. Mandates identifiers, keywords, etc.. in portable character set. Kido threatens appeal and throw-back to WD state.

“Localedef” can not produce the required format due to extended syntax. Keld: File format contains line end. Alain: code independent, interpretation by “compiler”. Kido: delete references to “file, byte, end of line concept”.

J36, J37 - accepted

J38 - accepted

J39 , J40 - accepted

Netherlands

NL1 - not limited to POSIX/UNIX/C, has been expanded to include a number of new issues.

NL1a - comment accepted, inappropriate for use

NL1b - outside the terms of reference for WG2

NL1c - working with liaison to WG15 about this issue.

NL2 - It is planned to add functionality, prioritized by available expertise and new requirements. Life expectancy of a standard is until it is withdrawn.

NL3a - Syntactical description is provided

NL3b - Changes have been made

NL3c - rejected, please identify.

NL3d - accepted, will be corrected.

NL3e - The conformance clause is being revised.

USA

4.1.1 - Properties of characters are not considered constant in 14652.

3.1.6 - FDCC will remain.

3.2.3 - Ellipses: Keld will check, possibly comply

4.1.2 - LC_COLLATE: rejected, more documentation is needed. We provide a customizable template, not an order correct for everybody and everything. Prehandling might be needed in certain cases.

C1 characters: presently not in 10646, might be some time.

Repertoiremap: combining acute accent - there are 2 such characters, both are included.

Discussion on mnemonic identifiers - Sato: this seems to be introducing a new character name. Winkler - can these be translated into 0646 identifiers? Keld: these identifiers are used in 40+ countries in specification of their locales. It would be very complicated to change them all. Kido: can we get the NBs that use them to express their preferences ? Keld: is also used in the internet. Sato: it seems that

this is a back way into the character identification again which will never go away, because it is perpetuated to support “old” locales.

Action item: Winkler to contact OpenGroup for permission to use their Layout APIs (page 45), and Layout Modifiers (page 39).

If the agreed upon dates can not be met, WG20 must consider the separation of relevant elements from 14652 for inclusion into 14651.

Future progression:

Proposal for time zones will go into the next version	Keld	Dec. 1, 1997
Produce draft for 1st FCD ballot	Keld	Dec. 1, 1997
Disposition of comments on web	Keld	December 1, 1997
Approval of DoC and FCD document	all	December 7, 1997
Send to Rinehuls for ballot	Winkler	December 16, 1997
SC22 FCD ballot in December 1997		before end of 1997
Receive ballot results		May 15, 1998
First draft of disposition of comments		May 22, 1998
Meeting		June 15, 1998

Acceptance criteria (after 1st FCD ballot): no new major functionality, no significant technical comments, no changes to conformance and scope (except “file”), text for standard must be included in the disposition of comments .

13. Internationalization API standard

536 Draft for ISO/IEC 15435 WD - I18N APIs	Keld Simonsen	97-10-04	22.15435
539 Issue list for ISO/IEC 15435 WD	Keld Simonsen	97-10-02	22.15435

Keld walks us through the documents. Keld has questions about the conformance statement - who or what will conform to this standard - implementers or other standards ???

Kido: liaison with WG11 seems to be necessary. Keld agrees to liaise with WG11 and to request that the API work be included in the WG11 standards.

Soor: it is needed to show the intention of the standard, especially also its adaptability to modern languages such as C++ and Java. Support for “text element”.

Kido: data type versus structure.

Soor: align chapters to the locales rather than the current structure of the document. Rename “message formatting” to “message handling”.

N539 - list of issues as seen by the editor: language independent, “abstract” character set, locale independent internal character representation.

Kido-Soor: tie issues to the proposed standard.

Winkler: coverage according to NP, not beyond, at least for the time being.

If Java or C++ binding is needed, the thread question must addressed.

Informative list of “old” vs “new” API for conversion in use in C programs.

Combining characters and directionality must be addressed - classification, sorting, equivalence, etc...

Linguistic word breaks, etc..

Use of new stuff - C++, Java, ...

Continuation of discussion on 11/20/97, based on issue list N554:

Language independent specification: no agreement - Canada does not accept WG11 standards and would have to vote against the standard, if mandated.

Bindings: enable various.

Character set: internal representation is implementation dependent. String data type, not tied to the encoding. Not locale dependent. no solution

Monetary: dual currency, eg. EURO, must be addressed.

Strings: termination by null byte or by length? We need to see the header .

Coverage: no comments

Thread-safe: implementation of operating system decides that. Operating on classes is so automatically.

Functionality: narrative, language independent, so that it can be implemented for any language.

Structure:

Rationale: why is this standard required? must be included in scope.

C++ binding: should be object oriented binding, good enough for Java or for C++

Word breaks, hyphenation: perhaps

Composite sequences: input needed.

Localedef API: Kido request!

How to check, if the API can be implemented ?

Spelling, dictionary, etc ...???

Action item Kung: suggest word break API

Action item Keld: look at Java and Win32 for complete coverage, talk to Mark Davis

New issue list N554

Progression:

develop new WD #2

14. ISO/IEC 10646 Issues

nothing to be done this time

15. Other business

15.1 Cultural adaptation workshop

535	CAW Information JTC1 N4627, N4789, N5018, GII-N123, SC2 and SC22 resolutions	Winkler - JTC1, SC2, SC22	97-09-21	admin
540	CAW: Preliminary agenda for the Cultural Adaptability Workshop in Ottawa, Jan. 20-22, 1998	Shibano	97-10-02	admin
541	ISO HLGS on CALS: Processing multiple languages, character sets and encoding	JTC1 N4851	97-07-28	admin
548	Contribution from WG20 to CAW - draft	SC22/WG20	97-11-15	admin

Arnold presented his proposed contribution to CAW as the basis for discussion within WG20.

Sato: our job is to find the place for WG20 in the new organization

Kido: confirm that the “new” WG20 would mainly be a advisory group, horizontally above the vertical product related groups.

Sato: we need to define, what WG20 work would be in the future. How can we increase the participation.

Khaled: can we change the scope of WG20 so that more people see the value of the work and participate

Sato: especially for I18N the voice of many people from many countries must be heard.

Keld: work is needed, this group is not particularly small - many other groups are same size. WG20 should stay in SC22 for successful continuation of their work.

Kido: danger that WG20 is killed due to only 4-5 NBs participating. We need to expand scope to attract more participants.

Khaled: how do we "market" our work.

Keld: we finally produce standards, now we attract people.

Baldev: interest in industry standards, modern technology. Little interest in legacy systems.

Kido: time to market is n issue. Standard itself does not sell, the binding does. WG20 members could participate in other groups as expert members.

Baldev: language bindings slow down the process, we are too late for current problems, e.g. EURO.

Arnold: Old technology is the major reason why Java people do not like POSIX.

Keld: we need to do the language bindings ourselves. Java is the same as we have been doing for a long time.

What does WG20 do today:

TR 11017
TR 10176
IS 14651
IS 14652
IS 15435

What do we know that WG20 could do:

euro
searching
character properties
special character use
rendition
input
spelling
hyphenation
address formats
telephone number format
keyboard, e-mail addresses
typography
rich text
time determination
holidays
calendary (1st day of the week, 1st working day, holidays, ...)
internet related searches
tagging

Looking at this list, there is nothing new as compared with the first meeting of WG20 in Cupertino. It is included in the TR 11017.

Non-known or undefined problems:

inter-networking communication
tagging tokens for cultural environments negotiation

Discussion on 11/17/97 does not give any agreement - the report will stop before any recommendation, just status of the current projects.

15.2 Discussion of EN 12005

Discussed on Thursday, 11/20/97.

Keld explains the registration standard, it is in fast track ballot - no comments and/or recommendations can be made as WG20. The approval of 14652 will require review of the registration standard.

15.3 SQL issue

Kido reports about SQL-3: only data type in binding is char. Octet interface must be used, possibly in UTF-8. Message to CAW that work in language bindings.

Action Winkler: contact Jim Melton, find TAG, request character to character bindings in SQL standard.

16. Review of Priorities and Target Dates

Reviewed.

17. Review of Actions Items from this meeting

Reviewed, no additions

18. Approval of Resolutions

All resolutions were approved without objection

19. Adjournment

The meeting was adjourned on November 20, 1997 at 12:15pm